(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 October 2001 (25.10.2001)

PCT

(10) International Publication Number
WO 01/79224 A2

(51) International Patent Classification[7]: C07H

(21) International Application Number: PCT/US01/11946

(22) International Filing Date: 12 April 2001 (12.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/196,408 12 April 2000 (12.04.2000) US

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BENTIVEGNA, Steven, C.** [US/US]; 49 Greenview Terrace, Middletown, CT 06457 (US). **CHEW, Anne** [US/US]; 1477 Beacon Street #64, Brookline, MA 02446 (US). **CHOI, Julie, Y.** [US/US]; 38 Elizabeth Street, West Haven, CT 06516 (US). **KOSHY, Beena** [IN/US]; Apt. 11B, 1298 Hartford Turnpike, North Haven, CT 06473 (US). **STEPHENS, J., Claiborne** [US/US]; 46 Crabapple Lane, Guilford, CT 06437 (US).

(74) Agent: **FIELD, Gisela, M.**; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/79224 A2

(54) Title: HAPLOTYPES OF THE CYP8B1 GENE

(57) Abstract: Novel single nucleotide polymorphisms in the human cytochrome P450 subfamily VIIIB (CYP8B1) gene are described. In addition, various genotypes, haplotypes and haplotype pairs for the CYP8B1 gene that exist in the population are described. Compositions and methods for haplotyping and/or genotyping the CYP8B1 gene in an individual are also disclosed. Polynucleotides containing one or more of the CYP8B1 polymorphisms disclosed herein are also described.

HAPLOTYPES OF THE CYP8B1 GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/196,408 filed April 12, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human cytochrome P450 subfamily VIIIB (CYP8B1) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the

population, leading to the failure of such drugs in clinical trials or their early withdrawal from the market even though they could be highly beneficial for other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* **15**:1249-52; Kleyn PW et al. 1998 *Science* **281**: 1820-21; Kola I 1999 *Curr Opin Biotech* **10**:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* **66**:445-7; Marshall, E 1999 *Science* **284**:406-7; Judson R et al. 2000 *Pharmacogenomics* **1**:1-12; Roses AD 2000 *Nature* **405**:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* **19**:216-7; Wang DG et al 1998 *Science* **280**:1077-82; Chakravarti A 1999 *Nat Genet* **21**:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* **4**:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* **5**:538-43; Davidson S 2000 *Nature Biotech* **18**:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* **161**: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* **320**:987-90; Dahl BS 1997 *Acta Psychiatr Scand* **96** (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* **63**:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* **97**:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of cardiovacular disorders is the cytochrome P450 subfamily VIIIB (CYP8B1) gene or its encoded product. CYP8B1, also known as sterol 12-alpha-hydroxylase, is an enzyme essential for the biosynthesis of cholic acid, a product of cholesterol metabolism. CYP8B1 determines the ratio of cholic acid (CA) to chenodeoxycholic acid

(CDCA), which in turn determines the hydrophobicity of bile acids. Both cholesterol levels and hydrophobicity of bile acids down-regulate the activity of CYP8B1. Thus, changes in the levels of cholesterol affect the activity of CYP8B1 which could be coupled to cardiovascular disorders associated with lipid metabolism (Vlahcevic et al., *Gastroenterology* 2000; 118:599-607).

Two DNA elements near the CYP8B1 promoter have been identified that are required for the promoter activity of CYP8B1. These two elements bind to Alpha(1)-fetoprotein transcription factor (FTF), a member of the nuclear receptor family, and regulate the activity of CYP8B1 promoter. Mutations in these two DNA elements suppress the CYP8B1 promoter activity. (Castillo-Olivares and Gil, *J. Biol. Chem.* 2000; 275:17793-17799). Also, CYP8B1 mRNA is regulated by thyroid hormone. This was concluded from studies in rats where thyroidectomy caused more than two-fold increase of CYP8B1 mRNA levels whereas treatment of intact rats with thyroxine resulted in a 50% reduction in mRNA levels (Andersson et al., *Biochim. Biophys. Acta* 1999; 1438:167-174). Thus, defects in the CYP8B1 gene are likely to result in cardiovascular disorders.

The cytochrome P450 subfamily VIIIB gene is located on chromosome 3p21.3-p22 and contains 1 exon that encodes a 501 amino acid protein. A reference sequence for the CYP8B1 gene is shown in Figure 1 (GenBank Accession No. AF090320.1; SEQ ID NO:1). Reference sequences for the coding sequence (GenBankAccession No. AF090320.1) and protein (Reference No. AAD19877.1) are shown in Figures 2 (SEQ ID NO:2) and 3 (SEQ ID NO:3), respectively.

Because of the potential for variation in the CYP8B1 gene to affect the expression and function of the encoded protein, it would be useful to know whether polymorphisms exist in the CYP8B1 gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of CYP8B1 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 9 novel polymorphic sites in the CYP8B1 gene. These polymorphic sites (PS) correspond to the following nucleotide positions in the indicated GenBank Accession Number: 1489 (PS1), 1671 (PS2), 1760 (PS3), 1946 (PS4), 2397 (PS5), 2482 (PS6), 2626 (PS7), 2753 (PS8) and 3115 (PS9) in AF090320.1. The polymorphisms at these sites are cytosine or thymine at PS1, guanine or adenine at PS2, cytosine or thymine at PS3, cytosine or thymine at PS4, adenine or guanine at PS5, guanine or thymine at PS6, guanine or adenine at PS7, cytosine or thymine at PS8 and guanine or adenine at PS9. In addition, the inventors have determined the identity of the alleles at these sites in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS9 in the CYP8B1 gene, which are shown below in Tables 5 and 4,

respectively. Each of these CYP8B1 haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the CYP8B1 gene that exists in the human population.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the CYP8B1 gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8 and PS9 in both copies of the CYP8B1 gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel CYP8B1 polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel CYP8B1 polymorphic sites. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 5 below or has one of the haplotype pairs in Table 4 below.

The invention also provides a method for haplotyping the CYP8B1 gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the CYP8B1 gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8 and PS9. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's CYP8B1 gene is defined by one of the CYP8B1 haplotypes shown in Table 5, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's CYP8B1 gene are defined by one of the CYP8B1 haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof. The method for establishing the CYP8B1 haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with CYP8B1 activity, e.g., cardiovacular disorders.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate CYP8B1 as a candidate target for treating a specific condition or disease predicted to be associated with CYP8B1 activity. Determining for a particular population the frequency of one or more of the individual CYP8B1 haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue CYP8B1 as a target for treating the specific disease of interest. In particular, if variable CYP8B1 activity is associated with the disease, then one or more CYP8B1 haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed CYP8B1 haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable CYP8B1 activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any CYP8B1 haplotype or haplotype pair, apply the information derived from detecting CYP8B1 haplotypes in an individual to decide whether modulating CYP8B1 activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting CYP8B1 to treat a specific condition or disease predicted to be associated with CYP8B1 activity. For example, detecting which of the CYP8B1 haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the most frequent CYP8B1 isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular CYP8B1 haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

The method for haplotyping the CYP8B1 gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with CYP8B1 activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the CYP8B1 haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute CYP8B1 haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of a CYP8B1 haplotype or haplotype pair that had a previously unknown association with response to the drug being studied in the trial. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any CYP8B1 haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a CYP8B1 genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the CYP8B1 genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the CYP8B1 genotype, haplotype, or haplotype pair in a reference population. A higher frequency of the CYP8B1 genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the CYP8B1 genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the CYP8B1 haplotype is selected from the haplotypes shown in Table 5, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for cardiovacular disorders.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the CYP8B1 gene or a fragment thereof. The reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at PS1, adenine at PS2, thymine at PS3, thymine at PS4, guanine at PS5, thymine at PS6, adenine at PS7, thymine at PS8 and adenine at PS9.

A particularly preferred polymorphic variant is an isogene of the CYP8B1 gene. A CYP8B1 isogene of the invention comprises cytosine or thymine at PS1, guanine or adenine at PS2, cytosine or thymine at PS3, cytosine or thymine at PS4, adenine or guanine at PS5, guanine or thymine at PS6, guanine or adenine at PS7, cytosine or thymine at PS8 and guanine or adenine at PS9. The invention also provides a collection of CYP8B1 isogenes, referred to herein as a CYP8B1 genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a CYP8B1 cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of thymine at a position corresponding to nucleotide 76, thymine at a position corresponding to nucleotide 262, guanine at a position corresponding to nucleotide 713, thymine at a position corresponding to nucleotide 798, adenine at a position corresponding to nucleotide 942, thymine at a position corresponding to nucleotide 1069 and adenine at a position corresponding to nucleotide 1431. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a CYP8B1 isogene defined by haplotypes 2-12.

Polynucleotides complementary to these CYP8B1 genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the CYP8B1-gene will be useful in studying the expression and function of CYP8B1, and in expressing CYP8B1 protein for use in screening for candidate drugs to treat diseases related to CYP8B1 activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express CYP8B1 for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the CYP8B1 protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of termination codon at a position corresponding to amino acid position 26, serine at a position corresponding to amino acid position 88, arginine at a position corresponding to amino acid position 238 and phenylalanine at a position corresponding to amino acid position 357. A polymorphic variant of CYP8B1 is useful in studying the effect of the variation on the biological activity of CYP8B1 as well as on the binding affinity of candidate drugs targeting CYP8B1 for the treatment of cardiovacular disorders.

The present invention also provides antibodies that recognize and bind to the above polymorphic CYP8B1 protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one of the CYP8B1 polymorphic genomic variants described herein and methods for producing such animals.

The transgenic animals are useful for studying expression of the CYP8B1 isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against CYP8B1 protein, and for testing the efficacy of therapeutic agents and compounds for cardiovacular disorders in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the CYP8B1 gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the CYP8B1 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing CYP8B1 haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the CYP8B1 gene (Genbank Accession Number AF090320.1; contiguous lines; SEQ ID NO:1), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:49 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R = G or A, Y = T or C, M = A or C, K = G or T, S = G or C, and W = A or T; WIPO standard ST.25).

Figure 2 illustrates a reference sequence for the CYP8B1 coding sequence (contiguous lines; SEQ ID NO:2) with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the CYP8B1 protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence. Any exclamation points (!) presented below the reference sequence represent a termination codon introduced by a polymorphism of Figure 2.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the CYP8B1 gene. As described in more detail below, the inventors herein discovered 12 isogenes of the CYP8B1 gene by characterizing the CYP8B1 gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (22 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (17 individuals). To the extent possible, the members of this reference population were organized into population subgroups by the self-identified

ethnogeographic origin of their four grandparents as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		22
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		17
	Caribbean	7
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The CYP8B1 isogenes present in the human reference population are defined by haplotypes for 9 polymorphic sites in the CYP8B1 gene, all of which are believed to be novel. The novel CYP8B1 polymorphic sites identified by the inventors are referred to as PS1-PS9 to designate the order in which they are located in the gene (see Table 3 below). Using the genotypes identified in the Index Repository for PS1-PS9 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the CYP8B1 gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the CYP8B1 gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether CYP8B1 is a suitable target for drugs to treat cardiovacular disorders, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular

nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all known polymorphic sites in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with

the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the CYP8B1 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel CYP8B1 polymorphisms and haplotypes identified herein.

The compositions comprise at least one CYP8B1 genotyping oligonucleotide. In one embodiment, a CYP8B1 genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a CYP8B1 polynucleotide, i.e., a CYP8B1 isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-CYP8B1 polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the CYP8B1 gene using the polymorphism information provided herein in conjunction with the known sequence information for the CYP8B1 gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention.

ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of

the polymorphic site to represent the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting CYP8B1 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

Accession No.: AF090320.1

AGTCAGCYAAGTGTT (SEQ ID NO:4) and its complement,
TTAATCCRCAGGAGC (SEQ ID NO:5) and its complement,
GATGCTCYGACAACG (SEQ ID NO:6) and its complement,
CTTTGGCYCCATCCT (SEQ ID NO:7) and its complement,
TTTCACARGATGCTC (SEQ ID NO:8) and its complement,
AGCAGGGKGTACCCT (SEQ ID NO:9) and its complement,
CTACCCARGTCCTGG (SEQ ID NO:10) and its complement,
CACCCTCYTCAGGTT (SEQ ID NO:11) and its complement, and
TTGACCCRCAGCCCT (SEQ ID NO:12) and its complement.

A preferred ASO primer for detecting CYP8B1 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

Accession No.:AF090320.1

GTGACCAGTCAGCYA (SEQ ID NO:13); GGACTTAACACTTRG (SEQ ID NO:14);
GAGAGCTTAATCCRC (SEQ ID NO:15); GGCTATGCTCCTGYG (SEQ ID NO:16);
GCCAGGGATGCTCYG (SEQ ID NO:17); GGCCTGCGTTGTCRG (SEQ ID NO:18);
CCTCTCCTTTGGCYC (SEQ ID NO:19); TCCTTGAGGATGGRG (SEQ ID NO:20);
CGTCTCTTTCACARG (SEQ ID NO:21); CACGGAGAGCATCYT (SEQ ID NO:22);
TGAGGGAGCAGGGKG (SEQ ID NO:23); TAGCTGAGGGTACMC (SEQ ID NO:24);
AGGAAGCTACCCARG (SEQ ID NO:25); CCTCACCCAGGACYT (SEQ ID NO:26);
TGCACCCACCCTCYT (SEQ ID NO:27); TGAACCAACCTGARG (SEQ ID NO:28);
CCCATGTTGACCCRC (SEQ ID NO:29); and AACCCCAGCGCTGYG (SEQ ID NO:30).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting CYP8B1 gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

Accession No.:AF090320.1

ACCAGTCAGC (SEQ ID NO:31); CTTAACACTT (SEQ ID NO:32);
AGCTTAATCC (SEQ ID NO:33); TATGCTCCTG (SEQ ID NO:34);
AGGGATGCTC (SEQ ID NO:35); CTGCGTTGTC (SEQ ID NO:36);
CTCCTTTGGC (SEQ ID NO:37); TTGAGGATGG (SEQ ID NO:38);

CTCTTTCACA (SEQ ID NO:39); GGAGAGCATC (SEQ ID NO:40);
GGGAGCAGGG (SEQ ID NO:41); CTGAGGGTAC (SEQ ID NO:42);
AAGCTACCCA (SEQ ID NO:43); CACCCAGGAC (SEQ ID NO:44);
ACCCACCCTC (SEQ ID NO:45); ACCAACCTGA (SEQ ID NO:46);
ATGTTGACCC (SEQ ID NO:47); and CCCAGCGCTG (SEQ ID NO:48).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

CYP8B1 genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized CYP8B1 genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the CYP8B1 gene in an individual. As used herein, the terms "CYP8B1 genotype" and "CYP8B1 haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the CYP8B1 gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid sample comprising the two copies of the CYP8B1 gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8 and PS9 in the two copies to assign a CYP8B1 genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a particularly preferred embodiment, the genotyping method comprises determining the identity of

the nucleotide pair at each of PS1-PS9.

Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the CYP8B1 gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ untranslated regions. If a CYP8B1 gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the CYP8B1 gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8 and PS9 in that copy to assign a CYP8B1 haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the CYP8B1 gene or fragment such as one of the methods described above for preparing CYP8B1 isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two CYP8B1 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional CYP8B1 clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the CYP8B1 gene in an individual. In a particularly preferred embodiment, the nucleotide at each of PS1-PS9 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the CYP8B1 haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual's CYP8B1 gene, the phased sequence of nucleotides present at each of PS1-PS9. The present invention also contemplates that typically only a subset of PS1-PS9 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 5. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, a CYP8B1 haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the

group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8 and PS9 in each copy of the CYP8B1 gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS9 in each copy of the CYP8B1 gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the CYP8B1 gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* **89**:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to

one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the CYP8B1 gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.*

17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's CYP8B1 haplotype pair is predicted from its CYP8B1 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a CYP8B1 genotype for the individual at two or more CYP8B1 polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing CYP8B1 haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the CYP8B1 haplotype pairs shown in Table 4.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African American, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to

$p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting a CYP8B1 haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*).

The invention also provides a method for determining the frequency of a CYP8B1 genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel CYP8B1 polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be a reference population, a family population, a same sex population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for CYP8B1 genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and a CYP8B1 genotype, haplotype, or haplotype pair. The trait may be

any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular CYP8B1 genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that CYP8B1 genotype, haplotype, or haplotype pair. Preferably, the CYP8B1 genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting CYP8B1 or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a CYP8B1 genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the

existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the CYP8B1 gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and CYP8B1 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their CYP8B1 genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the CYP8B1 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in PCT Application Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between CYP8B1 haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill,

New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra,* Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in PCT Application Serial No. PCT/US00/17540.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the CYP8B1 gene. As described in PCT Application Serial No. PCT/US00/17540, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of CYP8B1 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the CYP8B1 gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the CYP8B1 gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying CYP8B1 genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the CYP8B1 gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant CYP8B1 gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8 and PS9. Similarly, the nucleotide sequence of a variant fragment of the CYP8B1 gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the CYP8B1 gene, which is defined by haplotype 1, (or other reported CYP8B1 sequences) or to portions of the reference sequence (or other reported CYP8B1 sequences), except for genotyping oligonucleotides as described below.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of thymine

at PS1, adenine at PS2, thymine at PS3, thymine at PS4, guanine at PS5, thymine at PS6, adenine at PS7, thymine at PS8 and adenine at PS9. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the CYP8B1 gene which is defined by any one of haplotypes 2-12 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the CYP8B1 gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

CYP8B1 isogenes may be isolated using any method that allows separation of the two "copies" of the CYP8B1 gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides CYP8B1 genome anthologies, which are collections of CYP8B1 isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. A CYP8B1 genome anthology may comprise individual CYP8B1 isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the CYP8B1 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred CYP8B1 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded CYP8B1 protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct

combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant CYP8B1 sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the CYP8B1 gene will produce CYP8B1 mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a CYP8B1 cDNA comprising a nucleotide sequence which is a polymorphic variant of the CYP8B1 reference coding sequence shown in Figure 2. Thus, the invention also provides CYP8B1 mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO: 2 (Fig. 2), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 76, thymine at a position corresponding to nucleotide 262, guanine at a position corresponding to nucleotide 713, thymine at a position corresponding to nucleotide 798, adenine at a position corresponding to nucleotide 942, thymine at a position corresponding to nucleotide 1069 and adenine at a position corresponding to nucleotide 1431. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a CYP8B1 isogene defined by haplotypes 2-12. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized CYP8B1 cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of a CYP8B1 gene fragment comprises at least one

novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the CYP8B1 polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the CYP8B1 gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the CYP8B1 genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular CYP8B1 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the CYP8B1 isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular CYP8B1 isogene. Expression of a CYP8B1 isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions –10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of CYP8B1 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of CYP8B1 mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine

and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference CYP8B1 amino acid sequence shown in Figure 3. The location of a variant amino acid in a CYP8B1 polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig.3). A CYP8B1 protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 except for having one or more variant amino acids selected from the group consisting of termination codon at a position corresponding to amino acid position 26, serine at a position corresponding to amino acid position 88, arginine at a position corresponding to amino acid position 238 and phenylalanine at a position corresponding to amino acid position 357. The invention specifically excludes amino acid sequences identical to those previously identified for CYP8B1, including SEQ ID NO:3, and previously described fragments thereof. CYP8B1 protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, a CYP8B1 protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes shown in Table 5.

Table 2. Novel Polymorphic Variants of CYP8B1

Polymorphic Variant Number	Amino Acid Position and Identities			
	26	88	238	357
1	R	P	K	F
2	R	P	R	L
3	R	P	R	F
4	R	S	K	L
5	R	S	K	F
6	R	S	R	L
7	R	S	R	F
8	!			

The invention also includes CYP8B1 peptide variants, which are any fragments of a CYP8B1 protein variant that contain one or more of the amino acid variations shown in Table 2. A CYP8B1 peptide variant is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such CYP8B1 peptide variants may be useful as antigens to generate antibodies specific for one of the above CYP8B1 isoforms. In addition, the CYP8B1 peptide variants may be useful in drug screening assays.

A CYP8B1 variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant CYP8B1 genomic and cDNA sequences as described above. Alternatively, the CYP8B1 protein variant may be isolated from a biological sample of an individual having a CYP8B1 isogene which encodes the variant protein. Where the sample contains two

different CYP8B1 isoforms (i.e., the individual has different CYP8B1 isogenes), a particular CYP8B1 isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular CYP8B1 isoform but does not bind to the other CYP8B1 isoform.

The expressed or isolated CYP8B1 protein may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the CYP8B1 protein as discussed further below. CYP8B1 variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant CYP8B1 gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric CYP8B1 protein. The non-CYP8B1 portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the CYP8B1 and non-CYP8B1 portions so that the CYP8B1 protein may be cleaved and purified away from the non-CYP8B1 portion.

An additional embodiment of the invention relates to using a novel CYP8B1 protein isoform in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known CYP8B1 protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The CYP8B1 protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to a CYP8B1 variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the CYP8B1 protein(s) of interest and then washed. Bound CYP8B1 protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel CYP8B1 protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the CYP8B1 protein.

In yet another embodiment, when a particular CYP8B1 haplotype or group of CYP8B1 haplotypes encodes a CYP8B1 protein variant with an amino acid sequence distinct from that of CYP8B1 protein isoforms encoded by other CYP8B1 haplotypes, then detection of that particular CYP8B1 haplotype or group of CYP8B1 haplotypes may be accomplished by detecting expression of the encoded CYP8B1 protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive

with one or more of the novel CYP8B1 variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The CYP8B1 protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the CYP8B1 protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel protein isoforms described herein is administered to an individual to neutralize activity of the CYP8B1 isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the CYP8B1 protein variant from solution as well as react with CYP8B1 protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect CYP8B1 protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel CYP8B1 protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the CYP8B1 protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: Laboratory Techniques in Biochemistry and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci.USA 86;10029).

Effect(s) of the polymorphisms identified herein on expression of CYP8B1 may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the CYP8B1 gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into CYP8B1 protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired CYP8B1 isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the CYP8B1 isogene is introduced into a cell in such a way that it recombines with the endogenous CYP8B1 gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired CYP8B1 gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the CYP8B1 isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the CYP8B1 isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant CYP8B1 gene are prepared using standard procedures known in the art. Preferably, a construct comprising the

variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the CYP8B1 isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human CYP8B1 isogene and producing human CYP8B1 protein can be used as biological models for studying diseases related to abnormal CYP8B1 expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel CYP8B1 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel CYP8B1 isogenes; an antisense oligonucleotide directed against one of the novel CYP8B1 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel CYP8B1 isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel CYP8B1 isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art.

For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the CYP8B1 gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The CYP8B1 polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the CYP8B1 gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the CYP8B1 gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in the indicated GenBank Accession Number.

PCR Primer Pairs

GenBank Acc No.	Fragment No.	Forward Primer	Reverse Primer (complement of)	PCR Product
AF090320.1	Fragment 1	1116-1139	1732-1712	617 nt
AF090320.1	Fragment 2	1455-1476	2064-2041	610 nt
AF090320.1	Fragment 3	1480-1502	2150-2127	671 nt
AF090320.1	Fragment 4	1723-1744	2396-2375	674 nt
AF090320.1	Fragment 5	2010-2033	2588-2566	579 nt
AF090320.1	Fragment 6	2234-2255	2839-2819	606 nt
AF090320.1	Fragment 7	2564-2585	3197-3175	634 nt
AF090320.1	Fragment 8	2821-2841	3453-3431	633 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	= 6.6µl

Amplification profile:

97°C - 2 min. 1 cycle

97°C - 15 sec.
70°C - 45 sec.
72°C - 45 sec. } 10 cycles

97°C - 15 sec.
64°C - 45 sec.
72°C - 45 sec. } 35 cycles

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were

sequenced in both directions using the primer sets described previously or those represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in the indicated GenBank Accession Number. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

GenBank Acc No.	Fragment No.	Forward Primer	Reverse Primer (complement of)
AF090320.1	Fragment 1	1169-1188	1708-1689
AF090320.1	Fragment 2	1500-1519	2038-2018
AF090320.1	Fragment 3	1566-1584	2045-2026
AF090320.1	Fragment 4	1821-1841	2352-2333
AF090320.1	Fragment 5	2061-2080	2557-2538
AF090320.1	Fragment 6	2295-2315	2802-2783
AF090320.1	Fragment 7	2601-2621	3096-3077
AF090320.1	Fragment 8	2861-2880	3319-3300

Analysis of Sequences for Polymorphic Sites

Sequences were analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the CYP8B1 gene are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the CYP8B1 Gene

Polymorphic Site Number	PolyId[a]	Nucleotide Position	Reference Allele	Variant Allele
PS1	101495	1489(Acc#AF090320.1)	C	T
PS2	9719	1671(Acc#AF090320.1)	G	A
PS3	9720	1760(Acc#AF090320.1)	C	T
PS4	9723	1946(Acc#AF090320.1)	C	T
PS5	9725	2397(Acc#AF090320.1)	A	G
PS6	9726	2482(Acc#AF090320.1)	G	T
PS7	101501	2626(Acc#AF090320.1)	G	A
PS8	101502	2753(Acc#AF090320.1)	C	T
PS9	101521	3115(Acc#AF090320.1)	G	A

[a]PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.

EXAMPLE 2

This example illustrates analysis of the CYP8B1 polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred

based on linkage disequilibrium and/or Mendelian inheritance.

Table 4. Genotypes and Haplotype Pairs Observed for CYP8B1 Gene											
Genotype Number	Polymorphic Sites									HAP Pair	
	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9		
1	C	G	C	C	A	G	G	C	G	1	1
2	C	G	C	C	A	G	G	C	G/A	1	2
3	C	G	C	C	A	G	G	C/T	G	1	3
4	C	G	C	C	A/G	G	G	C	G	1	4
5	C	G	C	C/T	A	G	G	C	G	1	5
6	C	G	C	C	A	G/T	G	C	G/A	1	6
7	C	G	C	C	A/G	G/T	G	C	G/A	1	7
8	C	G	C	C	A/G	G	G	C/T	G	1	8
9	C	G/A	C	C	A	G	G	C	G	1	9
10	C/T	G	C	C	A	G	G	C	G	1	10
11	C	G	C	C	A	G	G/A	C	G	1	11
12	C	G	C/T	C	A	G	G	C	-	1	12
13	-	G	C	-	A	G	G	C	A	2	2

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family) and then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals.

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 12 human CYP8B1 haplotypes shown in Table 5 below.

Table 5. Haplotypes Identified in the CYP8B1 Gene									
Haplotype Number	Polymorphic Sites								
	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9
1	C	G	C	C	A	G	G	C	G
2	C	G	C	C	A	G	G	C	A
3	C	G	C	C	A	G	G	T	G
4	C	G	C	C	G	G	G	C	G
5	C	G	C	T	A	G	G	C	G
6	C	G	C	C	A	T	G	C	A
7	C	G	C	C	G	T	G	C	A
8	C	G	C	C	G	G	G	T	G
9	C	A	C	C	A	G	G	C	G
10	T	G	C	C	A	G	G	C	G
11	C	G	C	C	A	G	A	C	G
12	C	G	T	C	A	G	G	C	G

In Table 6 below, the number of chromosomes characterized by a given haplotype is shown, arranged by ethnic background of the subjects in the Index Repository. In Table 7 below, the number of subjects characterized by a given haplotype is shown, again arranged by ethnic background of the subjects in the Index Repository. In Tables 6 and 7, the following abbreviations are used: AF, African or African-American; AS, Asian; CA, Caucasian; HL, Hispanic-Latino; and AM, Native Americans.

Table 6. Frequencies of Observed Haplotypes in Non-Related Individuals

HAP No.	AF	AS	CA	HL	AM	Total
1	25	39	34	29	4	131
2	3	0	7	3	2	15
3	3	0	0	1	0	4
4	1	0	1	1	0	3
5	3	0	0	0	0	3
6	2	0	0	0	0	2
7	1	0	0	0	0	1
8	1	0	0	0	0	1
9	1	0	0	0	0	1
10	0	0	0	1	0	1
11	0	0	0	1	0	1
12	0	1	0	0	0	1

Table 6. Frequencies of Observed Haplotype Pairs

HAP Pair		AF	AS	CA	HL	AM	Total
1	1	5	19	20	11	1	56
2	1	3	0	5	3	2	13
2	2	0	0	1	0	0	1
3	1	8	0	0	1	0	9
4	1	1	0	1	1	0	3
5	1	3	0	0	0	0	3
6	1	2	0	0	0	0	2
7	1	1	0	0	0	0	1
8	1	1	0	0	0	0	1
9	1	1	0	0	0	0	1
10	1	0	0	0	1	0	1
11	1	0	0	0	1	0	1
12	1	0	1	0	0	0	1

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference

constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the cytochrome P450 subfamily VIIIB (CYP8B1) gene of an individual which comprises determining whether the individual has one of the CYP8B1 haplotypes shown in Table 5 or one of the haplotype pairs shown in Table 4.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS9 on at least one copy of the individual's CYP8B1 gene.

3. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS9 on both copies of the individual's CYP8B1 gene.

4. A method for genotyping the cytochrome P450 subfamily VIIIB (CYP8B1) gene of an individual, comprising determining for the two copies of the CYP8B1 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8 and PS9.

5. The method of claim 4, wherein the determining step comprises:
   (a) isolating from the individual a nucleic acid mixture comprising both copies of the CYP8B1 gene, or a fragment thereof, that are present in the individual;
   (b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;
   (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
   (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
   (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

6. The method of claim 4, which comprises determining for the two copies of the CYP8B1 gene present in the individual the identity of the nucleotide pair at each of PS1-PS9.

7. A method for haplotyping the cytochrome P450 subfamily VIIIB (CYP8B1) gene of an individual which comprises determining, for one copy of the CYP8B1 gene present in the individual, the identity of the nucleotide at two or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8 and PS9.

8. The method of claim 7, wherein the determining step comprises:
   (a) isolating from the individual a nucleic acid sample containing only one of the two copies of the CYP8B1 gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid molecule a target region containing the selected polymorphic site;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

9. A method for predicting a haplotype pair for the cytochrome P450 subfamily VIIIB (CYP8B1) gene of an individual comprising:

(a) identifying a CYP8B1 genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8 and PS9;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) comparing the possible haplotype pairs to the data in Table 4; and

(d) assigning a haplotype pair to the individual that is consistent with the data.

10. The method of claim 9, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS9.

11. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the cytochrome P450 subfamily VIIIB (CYP8B1) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-12 shown in Table 5 and the haplotype pair is selected from the haplotype pairs shown in Table 4, wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

12. The method of claim 11, wherein the trait is a clinical response to a drug targeting CYP8B1.

13. A composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the cytochrome P450 subfamily VIIIB (CYP8B1) gene at a polymorphic site selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8 and PS9.

14. The composition of claim 13, wherein the genotyping oligonucleotide is an allele-specific oligonucleotide that specifically hybridizes to an allele of the CYP8B1 gene at a region containing the polymorphic site.

15. The composition of claim 14, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-12, the complements

of SEQ ID NOS:4-12, and SEQ ID NOS:13-30.

16. The composition of claim 13, wherein the genotyping oligonucleotide is a primer-extension oligonucleotide.

17. The composition of claim 16, wherein the primer extension oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:31-48.

18. A kit for genotyping the CYP8B1 gene of an individual, which comprises a set of oligonucleotides designed to genotype each of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8 and PS9.

19. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which is a polymorphic variant of a reference sequence for the cytochrome P450 subfamily VIIIB (CYP8B1) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises a CYP8B1 isogene defined by a haplotype selected from the group consisting of haplotypes 1-12 in Table 5; and

(b) a second nucleotide sequence which is complementary to the first nucleotide sequence.

20. The isolated polynucleotide of claim 19, which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

21. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 19, wherein the organism expresses a CYP8B1 protein encoded by the first nucleotide sequence.

22. The recombinant organism of claim 21, which is a nonhuman transgenic animal.

23. The isolated polynucleotide of claim 19, wherein the first nucleotide sequence is a polymorphic variant of a fragment of the CYP8B1 gene, the fragment comprising one or more polymorphisms selected from the group consisting of thymine at PS1, adenine at PS2, thymine at PS3, thymine at PS4, guanine at PS5, thymine at PS6, adenine at PS7, thymine at PS8 and adenine at PS9.

24. An isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the CYP8B1 cDNA or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:2 and the polymorphic variant comprises the coding sequence of a CYP8B1 isogene defined by one of the haplotypes shown in Table 5.

25. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 24, wherein the organism expresses a cytochrome P450 subfamily VIIIB (CYP8B1) protein encoded by the polymorphic variant sequence.

26. The recombinant organism of claim 25, which is a nonhuman transgenic animal.

27. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the CYP8B1 protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:3 and the polymorphic variant is encoded by an isogene defined by one of the haplotypes shown in Table 5.

28. An isolated antibody specific for and immunoreactive with the isolated polypeptide of claim 27.

29. A method for screening for drugs targeting the isolated polypeptide of claim 27 which comprises contacting the CYP8B1 polymorphic variant with a candidate agent and assaying for binding activity.

30. A computer system for storing and analyzing polymorphism data for the cytochrome P450 subfamily VIIIB gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the genotypes and haplotype pairs shown in Table 4 and the haplotypes shown in Table 5.

31. A genome anthology for the cytochrome P450 subfamily VIIIB (CYP8B1) gene which comprises CYP8B1 isogenes defined by any one of haplotypes 1-12 shown in Table 5.

POLYMORPHISMS IN THE CYP8B1 GENE

```
GAATTCGTAA GGTTGAGGAG AGTTGATGAT GCCAAGTACT GTGGTGGTCC
TGAACAATTA AAGAGGGATT CTGGGAAGCA GAGTGTGGAC AGTTCCAACT      100
CCCTGCCAAG GGGAAGCTCA TAGGCAAAGG AAGCTCACTC CAGAGGGGAT
ATGGAAGTTC CATACCCTCT TTTGTCTGAA GAGCCGAAGT CCCTGTTCTC      200
AGGTCGTTAG GAAGTTAAAA AGTAATTTGG AGGTTATCAG AACTGATTGA
ATTGAGTTTG AACCTCACCT ATAGCAACAA TGGGCCAGGC TGCTTGACTA      300
ATGCCTTGGC GTCAATGGTA CAGTTTTCTC CCTCTTGAGC TGCTGGCAGG
GACCTGGGCT GACATGTCTC AGAAGGCCCT TAGTCAATGA TTAGCTTATC      400
TCAAGGCCCC AAGCCAGGGC AGCTGTCAAA GAGGGTCCCC ACTGCGTTCT
GCACCTAGAT CCTCATTGTG AAATGAAGTA TGAAGTGATT GAGTGAGGTC      500
TCTATTGTCT CTGACATTTT ACAATTCCAG GATTCTGCCT TCTTGTCAGA
GAAGTGTATA GGCAAGCAGT TGGGCAGGTG AGAGGGCTCC GGGTGAGAGG      600
CTCAGAGACT GAGGGCTCAG CCTCTGCTTG AAGAGTCATC ATCTGGGAGG
CTCTGTGGCC TCCTCAGATG AGTTCATTCA CACTCATACC CCAAAATGGA      700
GTCAACACCC CCTCCACCAT CTCAGCCTTC TCAGCATCTA AAGCCCCAGC
ATCGATGCCT CTTTTTTTGG GTTAGGGGTC AGAGCTGTTG TGGAAGGGCA      800
TACAGTCATT CTTCACTTGC CTTTGACTGT GTACTCTGTG CACATGGAGG
TAGGAGCAGA CATGACTTCA ACAAGGTCAT GCCCCCTTGG CAAGCATCTT      900
TGAGACCAGA GAGGAAGACA GACTAGGGAA AGAATGAGGA GATAAGCACG
GGCTGCTGTG AGGTCCAGGG GAGCAGGCAA AGGTAAGAGA AAAGGCTTTA     1000
GGATACTAAC TAACATATAT GGAGCACTAG CATGAGCCAG GCACTATTCT
AAGTGCTTTT CAGGTGTTAT CTCTTTTTGC CTCACGGACA GCACCTACAA     1100
GGCACTGTAA TTATCCCTAC TTCACAGATG AGGGAGTGGA GCCACAGTGA
GGTTAACTTA CTTGACCAAG GGGGCCAAGT AGGAATGGAG GCATTTGTTG     1200
AGTCTTCTAA AGATGAGGAA AGAGTGGAAG TGAGATTTTG TAAGTGCTTG
ATTCATTTCT ACCAACTGAA CTGGCAAATA AATAAAAGCA TGAGTAAATG     1300
GGGGTATAAA TAGTCTGTCA GCTATGGGGG TGGGAGTGGG CTCAAGGCAG
GCTTAGAGAG AAGGTGCAAG AGCTGTCTGA AAAGGTCAGA GCAAAGCATG     1400
AAGCTGGTGA GCAGCTGTGA CCATAGCTGG AAGCTTCTCT CTGAGCTTTC
TCCTGGTTAC CTCCTCCTCC CCTACGTGAC CAGTCAGCCA AGTGTTAAGT     1500
                                        T
CCAGGGGAAC ATTTTGCTGC TTCCAAGTAC TGTCTCACTA GTGTTATTTG
CCATAACTTG CGGCCACAGG GCAAGGTCCA GGTGCTCAGA CCTTTACATC     1600
CTGGACTTTC CAAGGCCTCC CAAAGCTCTC TGGCACCCAG GGAACAGTGT
GCGTGTCGAG AGCTTAATCC GCAGGAGCAT AGCCATGGTT CTCTGGGGTC     1700
                    A
      [EXON 1: 1685..
CAGTGCTGGG AGCTCTGCTG GTGGTCATTG CTGGATACCT GTGCCTGCCA
GGGATGCTCC GACAACGCAG GCCATGGGAG CCCCCTCTGG ACAAGGGTAC     1800
         T
CGTGCCCTGG CTTGGCCATG CCATGGCTTT CCGGAAGAAT ATGTTTGAAT
TTCTGAAGCG CATGAGGACC AAGCATGGGG ATGTGTTCAC AGTGCAGCTA     1900
GGGGGCCAGT ACTTCACCTT CGTCATGGAC CCCCTCTCCT TTGGCCCCAT
                                              T
CCTCAAGGAC ACACAGAGAA AACTAGACTT TGGGCAATAT GCAAAAAAAC     2000
TGGTGCTGAA GGTATTTGGA TACCGTTCAG TGCAAGGGGA CCATGAGATG
ATACACTCAG CCAGCACCAA GCATCTGAGG GGGGATGGCT TGAAGGATCT     2100
TAATGAGACC ATGCTGGACA GCCTGTCCTT TGTAATGCTG ACGTCCAAAG
GCTGGAGTCT GGATGCCAGT TGCTGGCATG AGGACAGCCT CTTTCGCTTC     2200
TGCTATTACA TCTTGTTCAC AGCTGGCTAC CTGAGCTTGT TCGGCTACAC
```

FIGURE 1A

```
GAAGGACAAG GAGCAGGACC TGCTACAGGC AGGAGAGTTA TTCATGGAGT      2300
TCCGCAAGTT TGACCTTCTT TTCCCAAGGT TTGTCTACTC CCTGCTGTGG
CCCCGGGAGT GGCTAGAAGT GGGCCGACTC CAGCGTCTCT TTCACAAGAT      2400
                                                    G
GCTCTCCGTG AGCCACAGCC AGGAGAAGGA GGGCATCAGC AACTGGCTGG
GCAACATGCT TCAGTTTCTG AGGGAGCAGG GGGTACCCTC AGCTATGCAG      2500
                                    T
GACAAGTTCA ACTTCATGAT GCTCTGGGCC TCCCAGGGGA ACACGGGGCC
TACCTCTTTC TGGGCCCTCT TGTACCTCCT GAAGCACCCA GAAGCTATTC      2600
GGGCTGTGAG GGAGGAAGCT ACCCAGGTCC TGGGTGAGGC CAGGCTGGAG
                          A
ACCAAGCAGT CCTTTGCCTT CAAACTCGGT GCCCTGCAAC ACACCCCAGT      2700
TCTAGACAGC GTGGTGGAGG AGACGCTGCG GCTGAGGGCT GCACCCACCC
TCCTCAGGTT GGTTCATGAA GACTATACCC TGAAGATGTC CAGTGGGCAG      2800
  T
GAGTATCTGT TCCGCCATGG AGACATCCTG GCCCTCTTTC CCTACCTCTC
AGTGCACATG GACCCTGACA TCCACCCTGA GCCCACCGTC TTCAAGTACG      2900
ATCGCTTCCT CAACCCTAAT GGCAGCCGGA AAGTGGACTT CTTCAAGACA
GGCAAGAAGA TCCACCACTA CACCATGCCC TGGGGTTCGG GCGTTTCCAT      3000
CTGCCCTGGG AGGTTCTTTG CACTCAGTGA GGTGAAGCTC TTTATCCTGC
TTATGGTCAC ACACTTTGAC TTAGAGTTGG TGGACCCTGA CACACCACTA      3100
CCCCATGTTG ACCCGCAGCG CTGGGGTTTT GGCACCATGC AGCCCAGCCA
               A
CGATGTGCGC TTCCGCTACC GCCTGCATCC TACAGAGTGA GCTTGGCCAA      3200
                 ..3190]
GCCAGCTGCA AACCTGGCCA GAGGAGTTCT ATTGCATCTC TCACCTGTTC
TCACCCCTCT GCAGCCCCAA GACCCCACTG GCCACCCCTC CCTCTGGTCC      3300
TGTGGCACCC CCTACCTCTG TTCTGCCTGT CCTCGCTCTC TCCCCGCCTA
GTCATCTGAC AGGCTTATCA TTCTCTTTAA AATACCATCT CTCAGAGTGG      3400
GTTCTGCCGA ACCCTCCTCT CACAGGAAGT CCAGAGGAAG GGGGAGTATC
TGTGGGCAAC TTGGTTTGGG AGATGATGCC TGCCTTGAGA AGTCCTGAGT      3500
ACAGAGACTG GTTCCCCCCA GACACGAGTA ACATGGCATC TTGCAAACAT
CAGCCTCCAC TCTCCCAGCT TGCTTTAGTT TTTTCAGCAA CACTTATCCC      3600
ACATCCTATG GAATTCAGGT TCTAGAACAG TGTCATCCAA CATAAATATG
AAGCAAGCTA CATGAGTAGT GGGGTTTTTT GGTTTTGTTT GTTTGTTTTG      3700
AGACAGAGTC TTGCTCTGTC GCCGAGGCTG CAGTGCAGCG GTGCGATCTC
TGCTCACTGC AACCTCTGCC TCCTGGGTTT AAGCAATTCT CCTGCCTCAG      3800
CCTCCCCAGT AGCTGGGATT ACAGGCACCT ACCACCTTGC CCAGCTAATT
TAGTTTCCTG GTAAACACAT TTTTAAAAAG TAAAATGAAA CAATTCATTT      3900
TTATAATATA CTTTACTTAA TCCAATATAT CCAAAATATT GGCATTTCAA
CGTGATCACT ATTAAAAATT TTAACGTAAT AGTTTTTACC TTCCCTTTTT      4000
CATCCTGTCT TCAAAGTCTG GTATGTACTG TACTTTTGCA GTTCTTCCCA
GTTCAGACTA GCCACATTCC AAGTGCTTAA TTGCCGTGTG TGGCAGGTGG      4100
CTGCCCTATT GGATACAACA GGTCTAGAGA AATGATACCT TTTTTTTTTT
TTTGAGACAG AGTCTCACTC TGTTGCCCAG GCTGCAGTGC AGTGGTGTGA      4200
TCTTGGCTCA CTACAACCTC TGCCTCCCGG GTTCAAGTAA TTCTCCTGCC
TCAGGCTTCT GAGTAGCTGG GATTATAGGT GCGCACCACC ACACCCGGCT      4300
AGTTTCTGTA TTTTTAGTAG AGACGGCGTT TCACCATTTT GGCCGGCCTC
CTCGGCCTCC CAAAGTTCTG GGATTACAGG CATGAGCCAC TGTGCCCAGT      4400
CAGGATGTCT TTAATGTAGG AATCATTCAA GATCCCTCCT CAGTGCCCAT
GTCTCCCCCA CCTCAGGGTG CTGTCAACCC TCCCTTGGTT TCCATGATCA      4500
TTGCTGAACT CAGAGCTTTC TTCTATCCCC AGACCCACAT GGGAGTCTCC
```

FIGURE 1B

```
CAGCACCTCT GACTCAGCTT GGCCAATGCA GAACTTGGCA TCTGAACCCC      4600
TGCCTCTTCC CTGAACCACT CCTCCTCCTG ATTTCCCTGC TCTCTCTGCT
CAGAACACTA CCATTCACCC CATCTCCCAG AGACACTCTC CCTCCTCCTC      4700
TTTCACCTTC ATACAAGCAG GCCCCCGTCC TATCAGTACA CCCCCCTGTG
ATGCCTCTGA AATATGTACC CTCCTCCCCT CTCCTACTGT CATTCTTATA      4800
CCTGTGATCT CTGCGTTTCT CTTCTACACT GCTGCCAGAG GCCTTTTTGC
CAAAAGCACT GCTGATTGTC ACTCTTTTAC TTTCCTGGTT CCCCTTTGAC      4900
TTCACACACT TGACGTTTAC ATTTTGTCAC AGTTGGTCCC CCTTGCCTTC
ACATACCTGA TGTTTAAATT TTGTCACGGT TGGTCTCCAA ACTTCTCAAC      5000
TTTGCCAGCT TTCTCCAATG AAACCATCAG TAGAGACATT GCTGTTCCTG
CCCATCCCAG TGCCAGCTAC TCCTTCCTGG AAGCCTTCCT GGATTTCACG      5100
CATCCCATCC TCTGGCCAGA AATGGCCTCC GAACTGTGCA GCTCCTGCGG
CACCATTTGT ATCATAGCAT ATCTGTTATG TGTCGTTGCT TTCTGTGTCT      5200
TGGTTTCTGC CTGCTGCCCA CTTCCCTAGA CTGGAGCCTA CTGGGGCATG
GTATTTTATT CATTTCTGAC CTCCAATTCC CTAAGTGGAG ATTAAATGTT      5300
AGAAACCTGA CTTCCAGTCT CGGCTAATTT ACTTTCAGTC CTCAAATCAT
TTCCTGGCCT CAGCCCTCCA ATCTATGAAA TGAGGACAAT CCCCCTTCCC      5400
AGCGTGCTTA TCAGTGCTAC AAGAGGGTGA GGGGGTGGCT GCCATAGCTG
CAGCTGGCTG ATGGCAGGGC TCCAGCTGGA GCTGGGACAG GAGGAAGAAT      5500
GACTG                                                       5505
```

FIGURE 1C

POLYMORPHISMS IN THE CODING SEQUENCE OF CYP8B1

```
ATGGTTCTCT GGGGTCCAGT GCTGGGAGCT CTGCTGGTGG TCATTGCTGG
ATACCTGTGC CTGCCAGGGA TGCTCCGACA ACGCAGGCCA TGGGAGCCCC      100
                          T
CTCTGGACAA GGGTACCGTG CCCTGGCTTG GCCATGCCAT GGCTTTCCGG
AAGAATATGT TTGAATTTCT GAAGCGCATG AGGACCAAGC ATGGGGATGT      200
GTTCACAGTG CAGCTAGGGG GCCAGTACTT CACCTTCGTC ATGGACCCCC
TCTCCTTTGG CCCCATCCTC AAGGACACAC AGAGAAAACT AGACTTTGGG      300
    T
CAATATGCAA AAAAACTGGT GCTGAAGGTA TTTGGATACC GTTCAGTGCA
AGGGGACCAT GAGATGATAC ACTCAGCCAG CACCAAGCAT CTGAGGGGGG      400
ATGGCTTGAA GGATCTTAAT GAGACCATGC TGGACAGCCT GTCCTTTGTA
ATGCTGACGT CCAAAGGCTG GAGTCTGGAT GCCAGTTGCT GGCATGAGGA      500
CAGCCTCTTT CGCTTCTGCT ATTACATCTT GTTCACAGCT GGCTACCTGA
GCTTGTTCGG CTACACGAAG GACAAGGAGC AGGACCTGCT ACAGGCAGGA      600
GAGTTATTCA TGGAGTTCCG CAAGTTTGAC CTTCTTTTCC CAAGGTTTGT
CTACTCCCTG CTGTGGCCCC GGGAGTGGCT AGAAGTGGGC CGACTCCAGC      700
GTCTCTTTCA CAAGATGCTC TCCGTGAGCC ACAGCCAGGA GAAGGAGGGC
            G
ATCAGCAACT GGCTGGGCAA CATGCTTCAG TTTCTGAGGG AGCAGGGGGT      800
                                                    T
ACCCTCAGCT ATGCAGGACA AGTTCAACTT CATGATGCTC TGGGCCTCCC
AGGGGAACAC GGGGCCTACC TCTTTCTGGG CCCTCTTGTA CCTCCTGAAG      900
CACCCAGAAG CTATTCGGGC TGTGAGGGAG GAAGCTACCC AGGTCCTGGG
                                          A
TGAGGCCAGG CTGGAGACCA AGCAGTCCTT TGCCTTCAAA CTCGGTGCCC     1000
TGCAACACAC CCCAGTTCTA GACAGCGTGG TGGAGGAGAC GCTGCGGCTG
AGGGCTGCAC CCACCCTCCT CAGGTTGGTT CATGAAGACT ATACCCTGAA     1100
                T
GATGTCCAGT GGGCAGGAGT ATCTGTTCCG CCATGGAGAC ATCCTGGCCC
TCTTTCCCTA CCTCTCAGTG CACATGGACC CTGACATCCA CCCTGAGCCC     1200
ACCGTCTTCA AGTACGATCG CTTCCTCAAC CCTAATGGCA GCCGGAAAGT
GGACTTCTTC AAGACAGGCA AGAAGATCCA CCACTACACC ATGCCCTGGG     1300
GTTCGGGCGT TTCCATCTGC CCTGGGAGGT TCTTTGCACT CAGTGAGGTG
AAGCTCTTTA TCCTGCTTAT GGTCACACAC TTTGACTTAG AGTTGGTGGA     1400
CCCTGACACA CCACTACCCC ATGTTGACCC GCAGCGCTGG GGTTTTGGCA
                              A
CCATGCAGCC CAGCCACGAT GTGCGCTTCC GCTACCGCCT GCATCCTACA     1500
GAGTGA                                                    1506
```

FIGURE 2

ISOFORMS OF THE CYP8B1 PROTEIN

```
MVLWGPVLGA LLVVIAGYLC LPGMLRQRRP WEPPLDKGTV PWLGHAMAFR
                          !
KNMFEFLKRM RTKHGDVFTV QLGGQYFTFV MDPLSFGPIL KDTQRKLDFG      100
                                    S
QYAKKLVLKV FGYRSVQGDH EMIHSASTKH LRGDGLKDLN ETMLDSLSFV
MLTSKGWSLD ASCWHEDSLF RFCYYILFTA GYLSLFGYTK DKEQDLLQAG      200
ELFMEFRKFD LLFPRFVYSL LWPREWLEVG RLQRLFHKML SVSHSQEKEG
                                  R
ISNWLGNMLQ FLREQGVPSA MQDKFNFMML WASQGNTGPT SFWALLYLLK      300
HPEAIRAVRE EATQVLGEAR LETKQSFAFK LGALQHTPVL DSVVEETLRL
RAAPTLLRLV HEDYTLKMSS GQEYLFRHGD ILALFPYLSV HMDPDIHPEP      400
      F
TVFKYDRFLN PNGSRKVDFF KTGKKIHHYT MPWGSGVSIC PGRFFALSEV
KLFILLMVTH FDLELVDPDT PLPHVDPQRW GFGTMQPSHD VRFRYRLHPT      500
E                                                           501
```

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Bentivegna, Steven C.
Chew, Anne
Choi, Julie Y.
Koshy, Beena
Stephens, J. Claiborne

<120> Haplotypes of the CYBP8B1 Gene

<130> MWH-0400PCT CYP8B1

<140> To be assigned
<141> 2001-04-12

<150> 60/196,408
<151> 2000-04-12

<160> 49

<170> PatentIn Ver. 2.1

<210> 1
<211> 5505
<212> DNA
<213> Homo sapien

<400> 1
gaattcgtaa ggttgaggag agttgatgat gccaagtact gtggtggtcc tgaacaatta 60
aagagggatt ctgggaagca gagtgtggac agttccaact ccctgccaag gggaagctca 120
taggcaaagg aagctcactc cagaggggat atggaagttc catacccctct tttgtctgaa 180
gagccgaagt ccctgttctc aggtcgttag gaagttaaaa agtaatttgg aggttatcag 240
aactgattga attgagtttg aacctcacct atagcaacaa tgggccaggc tgcttgacta 300
atgccttggc gtcaatggta cagttttctc cctcttgagc tgctggcagg gacctgggct 360
gacatgtctc agaaggccct tagtcaatga ttagcttatc tcaaggcccc aagccagggc 420
agctgtcaaa gagggtcccc actgcgttct gcacctagat cctcattgtg aaatgaagta 480
tgaagtgatt gagtgaggtc tctattgtct ctgacatttt acaattccag gattctgcct 540
tcttgtcaga gaagtgtata ggcaagcagt tgggcaggtg agagggctcc gggtgagagg 600
ctcagagact gagggctcag cctctgcttg aagagtcatc atctgggagg ctctgtggcc 660
tcctcagatg agttcattca cactcatacc ccaaaatgga gtcaacaccc cctccaccat 720
ctcagccttc tcagcatcta aagccccagc atcgatgcct ctttttttgg gttaggggtc 780
agagctgttg tggaagggca tacagtcatt cttcacttgc ctttgactgt gtactctgtg 840
cacatggagg taggagcaga catgacttca acaaggtcat gccccccttgg caagcatctt 900
tgagaccaga gaggaagaca gactagggaa agaatgagga gataagcacg ggctgctgtg 960
aggtccaggg gagcaggcaa aggtaagaga aaaggcttta ggatactaac taacatatat 1020
ggagcactag catgagccag gcactattct aagtgctttt caggtgttat ctctttttgc 1080
ctcacggaca gcacctacaa ggcactgtaa ttatccctac ttcacagatg agggagtgga 1140

gccacagtga ggttaactta cttgaccaag ggggccaagt aggaatggag gcatttgttg 1200
agtcttctaa agatgaggaa agagtggaag tgagattttg taagtgcttg attcatttct 1260
accaactgaa ctggcaaata aataaaagca tgagtaaatg ggggtataaa tagtctgtca 1320
gctatggggg tgggagtggg ctcaaggcag gcttagagag aaggtgcaag agctgtctga 1380
aaaggtcaga gcaaagcatg aagctggtga gcagctgtga ccatagctgg aagcttctct 1440
ctgagctttc tcctggttac ctcctcctcc cctacgtgac cagtcagcca agtgttaagt 1500
ccaggggaac attttgctgc ttccaagtac tgtctcacta gtgttatttg ccataacttg 1560
cggccacagg gcaaggtcca ggtgctcaga cctttacatc ctggactttc caaggcctcc 1620
caaagctctc tggcacccag ggaacagtgt gcgtgtcgag agcttaatcc gcaggagcat 1680
agccatggtt ctctggggtc cagtgctggg agctctgctg gtggtcattg ctggatacct 1740
gtgcctgcca gggatgctcc gacaacgcag gccatgggag ccccctctgg acaagggtac 1800
cgtgccctgg cttggccatg ccatggcttt ccggaagaat atgtttgaat ttctgaagcg 1860
catgaggacc aagcatgggg atgtgttcac agtgcagcta gggggccagt acttcacctt 1920
cgtcatggac cccctctcct ttggccccat cctcaaggac acacagagaa aactagactt 1980
tgggcaatat gcaaaaaaac tggtgctgaa ggtatttgga taccgttcag tgcaagggga 2040
ccatgagatg atacactcag ccagcaccaa gcatctgagg ggggatggct tgaaggatct 2100
taatgagacc atgctggaca gcctgtcctt tgtaatgctg acgtccaaag gctggagtct 2160
ggatgccagt tgctggcatg aggacagcct ctttcgcttc tgctattaca tcttgttcac 2220
agctggctac ctgagcttgt tcggctacac gaaggacaag gagcaggacc tgctacaggc 2280
aggagagtta ttcatggagt tccgcaagtt tgaccttctt ttcccaaggt ttgtctactc 2340
cctgctgtgg ccccgggagt ggctagaagt gggccgactc cagcgtctct ttcacaagat 2400
gctctccgtg agccacagcc aggagaagga gggcatcagc aactggctgg gcaacatgct 2460
tcagtttctg agggagcagg gggtaccctc agctatgcag gacaagttca acttcatgat 2520
gctctgggcc tcccagggga acacggggcc tacctctttc tgggccctct tgtacctcct 2580
gaagcaccca gaagctattc gggctgtgag ggaggaagct acccaggtcc tgggtgaggc 2640
caggctggag accaagcagt cctttgcctt caaactcggt gccctgcaac acaccccagt 2700
tctagacagc gtggtggagg agacgctgcg gctgagggct gcacccaccc tcctcaggtt 2760
ggttcatgaa gactataccc tgaagatgtc cagtgggcag gagtatctgt tccgccatgg 2820
agacatcctg gccctctttc cctacctctc agtgcacatg gaccctgaca tccaccctga 2880
gcccaccgtc ttcaagtacg atcgcttcct caaccctaat ggcagccgga aagtggactt 2940
cttcaagaca ggcaagaaga tccaccacta caccatgccc tggggttcgg gcgtttccat 3000
ctgccctggg aggttctttg cactcagtga ggtgaagctc tttatcctgc ttatggtcac 3060
acactttgac ttagagttgg tggaccctga cacaccacta ccccatgttg acccgcagcg 3120
ctggggtttt ggcaccatgc agcccagcca cgatgtgcgc ttccgctacc gcctgcatcc 3180
tacagagtga gcttggccaa gccagctgca aacctggcca gaggagttct attgcatctc 3240
tcacctgttc tcacccctct gcagccccaa gaccccactg gccacccctc cctctggtcc 3300
tgtggcaccc cctacctctg ttctgcctgt cctcgctctc tccccgccta gtcatctgac 3360
aggcttatca ttctctttaa aataccatct ctcagagtgg gttctgccga accctcctct 3420
cacaggaagt ccagaggaag ggggagtatc tgtgggcaac ttggtttggg agatgatgcc 3480
tgccttgaga agtcctgagt acagagactg gttcccccca gacacgagta acatggcatc 3540
ttgcaaacat cagcctccac tctcccagct tgctttagtt ttttcagcaa cacttatccc 3600
acatcctatg gaattcaggt tctagaacag tgtcatccaa cataaatatg aagcaagcta 3660
catgagtagt ggggtttttt ggttttgttt gtttgttttg agacagagtc ttgctctgtc 3720
gccgaggctg cagtgcagcg gtgcgatctc tgctcactgc aacctctgcc tcctgggttt 3780
aagcaattct cctgcctcag cctccccagt agctgggatt acaggcacct accaccttgc 3840
ccagctaatt tagtttcctg gtaaacacat ttttaaaaag taaaatgaaa caattcattt 3900
ttataatata ctttacttaa tccaatatat ccaaaatatt ggcatttcaa cgtgatcact 3960
attaaaaatt ttaacgtaat agtttttacc ttccctttt catcctgtct tcaaagtctg 4020

gtatgtactg tacttttgca gttcttccca gttcagacta gccacattcc aagtgcttaa 4080
ttgccgtgtg tggcaggtgg ctgccctatt ggatacaaca ggtctagaga aatgatacct 4140
ttttttttt tttgagacag agtctcactc tgttgcccag gctgcagtgc agtggtgtga 4200
tcttggctca ctacaacctc tgcctcccgg gttcaagtaa ttctcctgcc tcaggcttct 4260
gagtagctgg gattataggt gcgcaccacc acacccggct agtttctgta tttttagtag 4320
agacggcgtt tcaccatttt ggccggcctc ctcggcctcc caaagttctg ggattacagg 4380
catgagccac tgtgcccagt caggatgtct ttaatgtagg aatcattcaa gatccctcct 4440
cagtgcccat gtctccccca cctcagggtg ctgtcaaccc tcccttggtt tccatgatca 4500
ttgctgaact cagagctttc ttctatcccc agacccacat gggagtctcc cagcacctct 4560
gactcagctt ggccaatgca gaacttggca tctgaacccc tgcctcttcc ctgaaccact 4620
cctcctcctg atttccctgc tctctctgct cagaacacta ccattcaccc catctcccag 4680
agacactctc cctcctcctc tttcaccttc atacaagcag gcccccgtcc tatcagtaca 4740
ccccccctgtg atgcctctga aatatgtacc ctcctccct ctcctactgt cattcttata 4800
cctgtgatct ctgcgtttct cttctacact gctgccagag gccttttgc caaaagcact 4860
gctgattgtc actcttttac tttcctggtt ccccctttgac ttcacacact tgacgtttac 4920
attttgtcac agttggtccc ccttgccttc acatacctga tgtttaaatt ttgtcacggt 4980
tggtctccaa acttctcaac tttgccagct ttctccaatg aaaccatcag tagagacatt 5040
gctgttcctg cccatcccag tgccagctac tccttcctgg aagccttcct ggatttcacg 5100
catcccatcc tctggccaga aatggcctcc gaactgtgca gctcctgcgg caccatttgt 5160
atcatagcat atctgttatg tgtcgttgct ttctgtgtct tggtttctgc ctgctgccca 5220
cttccctaga ctggagccta ctggggcatg gtattttatt catttctgac ctccaattcc 5280
ctaagtggag attaaatgtt agaaacctga cttccagtct cggctaattt actttcagtc 5340
ctcaaatcat ttcctggcct cagccctcca atctatgaaa tgaggacaat ccccttcccc 5400
agcgtgctta tcagtgctac aagagggtga gggggtggct gccatagctg cagctggctg 5460
atggcagggc tccagctgga gctgggacag gaggaagaat gactg 5505

<210> 2
<211> 1506
<212> DNA
<213> Homo sapien

<400> 2
atggttctct ggggtccagt gctgggagct ctgctggtgg tcattgctgg atacctgtgc 60
ctgccaggga tgctccgaca acgcaggcca tgggagcccc ctctggacaa gggtaccgtg 120
ccctggcttg gccatgccat ggctttccgg aagaatatgt ttgaatttct gaagcgcatg 180
aggaccaagc atggggatgt gttcacagtg cagctagggg gccagtactt caccttcgtc 240
atggaccccc tctcctttgg ccccatcctc aaggacacac agagaaaact agactttggg 300
caatatgcaa aaaaactggt gctgaaggta tttggatacc gttcagtgca aggggaccat 360
gagatgatac actcagccag caccaagcat ctgaggggggg atggcttgaa ggatcttaat 420
gagaccatgc tggacagcct gtcctttgta atgctgacgt ccaaaggctg gagtctggat 480
gccagttgct ggcatgagga cagcctcttt cgcttctgct attacatctt gttcacagct 540
ggctacctga gcttgttcgg ctacacgaag gacaaggagc aggacctgct acaggcagga 600
gagttattca tggagttccg caagtttgac cttcttttcc caaggtttgt ctactccctg 660
ctgtggcccc gggagtggct agaagtgggc cgactccagc gtctctttca caagatgctc 720
tccgtgagcc acagccagga gaaggagggc atcagcaact ggctgggcaa catgcttcag 780
tttctgaggg agcaggggggt accctcagct atgcaggaca agttcaactt catgatgctc 840
tgggcctccc aggggaacac ggggcctacc tctttctggg ccctcttgta cctcctgaag 900

cacccagaag ctattcgggc tgtgagggag gaagctaccc aggtcctggg tgaggccagg 960
ctggagacca agcagtcctt tgccttcaaa ctcggtgccc tgcaacacac cccagttcta 1020
gacagcgtgg tggaggagac gctgcggctg agggctgcac ccaccctcct caggttggtt 1080
catgaagact ataccctgaa gatgtccagt gggcaggagt atctgttccg ccatggagac 1140
atcctggccc tctttcccta cctctcagtg cacatggacc ctgacatcca ccctgagccc 1200
accgtcttca agtacgatcg cttcctcaac cctaatggca gccggaaagt ggacttcttc 1260
aagacaggca agaagatcca ccactacacc atgccctggg gttcgggcgt ttccatctgc 1320
cctgggaggt tctttgcact cagtgaggtg aagctcttta tcctgcttat ggtcacacac 1380
tttgacttag agttggtgga ccctgacaca ccactacccc atgttgaccc gcagcgctgg 1440
ggttttggca ccatgcagcc cagccacgat gtgcgcttcc gctaccgcct gcatcctaca 1500
gagtga 1506

<210> 3
<211> 501
<212> PRT
<213> Homo sapien

<400> 3
Met Val Leu Trp Gly Pro Val Leu Gly Ala Leu Leu Val Val Ile Ala
1 5 10 15

Gly Tyr Leu Cys Leu Pro Gly Met Leu Arg Gln Arg Arg Pro Trp Glu
20 25 30

Pro Pro Leu Asp Lys Gly Thr Val Pro Trp Leu Gly His Ala Met Ala
35 40 45

Phe Arg Lys Asn Met Phe Glu Phe Leu Lys Arg Met Arg Thr Lys His
50 55 60

Gly Asp Val Phe Thr Val Gln Leu Gly Gly Gln Tyr Phe Thr Phe Val
65 70 75 80

Met Asp Pro Leu Ser Phe Gly Pro Ile Leu Lys Asp Thr Gln Arg Lys
85 90 95

Leu Asp Phe Gly Gln Tyr Ala Lys Lys Leu Val Leu Lys Val Phe Gly
100 105 110

Tyr Arg Ser Val Gln Gly Asp His Glu Met Ile His Ser Ala Ser Thr
115 120 125

Lys His Leu Arg Gly Asp Gly Leu Lys Asp Leu Asn Glu Thr Met Leu
130 135 140

Asp Ser Leu Ser Phe Val Met Leu Thr Ser Lys Gly Trp Ser Leu Asp
145 150 155 160

```
Ala Ser Cys Trp His Glu Asp Ser Leu Phe Arg Phe Cys Tyr Tyr Ile
                165                 170                 175

Leu Phe Thr Ala Gly Tyr Leu Ser Leu Phe Gly Tyr Thr Lys Asp Lys
            180                 185                 190

Glu Gln Asp Leu Leu Gln Ala Gly Glu Leu Phe Met Glu Phe Arg Lys
        195                 200                 205

Phe Asp Leu Leu Phe Pro Arg Phe Val Tyr Ser Leu Leu Trp Pro Arg
    210                 215                 220

Glu Trp Leu Glu Val Gly Arg Leu Gln Arg Leu Phe His Lys Met Leu
225                 230                 235                 240

Ser Val Ser His Ser Gln Glu Lys Glu Gly Ile Ser Asn Trp Leu Gly
                245                 250                 255

Asn Met Leu Gln Phe Leu Arg Glu Gln Gly Val Pro Ser Ala Met Gln
            260                 265                 270

Asp Lys Phe Asn Phe Met Met Leu Trp Ala Ser Gln Gly Asn Thr Gly
        275                 280                 285

Pro Thr Ser Phe Trp Ala Leu Leu Tyr Leu Leu Lys His Pro Glu Ala
    290                 295                 300

Ile Arg Ala Val Arg Glu Glu Ala Thr Gln Val Leu Gly Glu Ala Arg
305                 310                 315                 320

Leu Glu Thr Lys Gln Ser Phe Ala Phe Lys Leu Gly Ala Leu Gln His
                325                 330                 335

Thr Pro Val Leu Asp Ser Val Val Glu Glu Thr Leu Arg Leu Arg Ala
            340                 345                 350

Ala Pro Thr Leu Leu Arg Leu Val His Glu Asp Tyr Thr Leu Lys Met
        355                 360                 365

Ser Ser Gly Gln Glu Tyr Leu Phe Arg His Gly Asp Ile Leu Ala Leu
    370                 375                 380

Phe Pro Tyr Leu Ser Val His Met Asp Pro Asp Ile His Pro Glu Pro
385                 390                 395                 400

Thr Val Phe Lys Tyr Asp Arg Phe Leu Asn Pro Asn Gly Ser Arg Lys
                405                 410                 415
```

Val Asp Phe Phe Lys Thr Gly Lys Lys Ile His His Tyr Thr Met Pro
420 425 430

Trp Gly Ser Gly Val Ser Ile Cys Pro Gly Arg Phe Phe Ala Leu Ser
435 440 445

Glu Val Lys Leu Phe Ile Leu Leu Met Val Thr His Phe Asp Leu Glu
450 455 460

Leu Val Asp Pro Asp Thr Pro Leu Pro His Val Asp Pro Gln Arg Trp
465 470 475 480

Gly Phe Gly Thr Met Gln Pro Ser His Asp Val Arg Phe Arg Tyr Arg
485 490 495

Leu His Pro Thr Glu
500

<210> 4
<211> 15
<212> DNA
<213> Homo sapien

<400> 4
agtcagcyaa gtgtt 15

<210> 5
<211> 15
<212> DNA
<213> Homo sapien

<400> 5
ttaatccrca ggagc 15

<210> 6
<211> 15
<212> DNA
<213> Homo sapien

<400> 6
gatgctcyga caacg 15

<210> 7

<211> 15
<212> DNA
<213> Homo sapien

<400> 7
ctttggcycc atcct 15

<210> 8
<211> 15
<212> DNA
<213> Homo sapien

<400> 8
tttcacarga tgctc 15

<210> 9
<211> 15
<212> DNA
<213> Homo sapien

<400> 9
agcagggkgt accct 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapien

<400> 10
ctacccargt cctgg 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapien

<400> 11
caccctcytc aggtt 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapien

<400> 12
ttgacccrca gcgct 15

<210> 13
<211> 15
<212> DNA
<213> Homo sapien

<400> 13
gtgaccagtc agcya 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapien

<400> 14
ggacttaaca cttrg 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapien

<400> 15
gagagcttaa tccrc 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapien

<400> 16
ggctatgctc ctgyg 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapien

<400> 17
gccagggatg ctcyg 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapien

<400> 18
ggcctgcgtt gtcrg 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapien

<400> 19
cctctccttt ggcyc 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapien

<400> 20
tccttgagga tggrg 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapien

<400> 21
cgtctctttc acarg 15

<210> 22
<211> 15
<212> DNA
<213> Homo sapien

<400> 22
cacggagagc atcyt 15

<210> 23

<211> 15
<212> DNA
<213> Homo sapien

<400> 23
tgagggagca gggkg 15

<210> 24
<211> 15
<212> DNA
<213> Homo sapien

<400> 24
tagctgaggg tacmc 15

<210> 25
<211> 15
<212> DNA
<213> Homo sapien

<400> 25
aggaagctac ccarg 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapien

<400> 26
cctcacccag gacyt 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapien

<400> 27
tgcacccacc ctcyt 15

<210> 28
<211> 15
<212> DNA
<213> Homo sapien

<400> 28
tgaaccaacc tgarg 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapien

<400> 29
cccatgttga cccrc 15

<210> 30
<211> 15
<212> DNA
<213> Homo sapien

<400> 30
aaccccagcg ctgyg 15

<210> 31
<211> 10
<212> DNA
<213> Homo sapien

<400> 31
accagtcagc 10

<210> 32
<211> 10
<212> DNA
<213> Homo sapien

<400> 32
cttaacactt 10

<210> 33
<211> 10
<212> DNA
<213> Homo sapien

<400> 33
agcttaatcc 10

<210> 34
<211> 10
<212> DNA
<213> Homo sapien

<400> 34
tatgctcctg 10

<210> 35
<211> 10
<212> DNA
<213> Homo sapien

<400> 35
agggatgctc 10

<210> 36
<211> 10
<212> DNA
<213> Homo sapien

<400> 36
ctgcgttgtc 10

<210> 37
<211> 10
<212> DNA
<213> Homo sapien

<400> 37
ctcctttggc 10

<210> 38
<211> 10
<212> DNA
<213> Homo sapien

<400> 38
ttgaggatgg 10

<210> 39

<211> 10
<212> DNA
<213> Homo sapien

<400> 39
ctctttcaca 10

<210> 40
<211> 10
<212> DNA
<213> Homo sapien

<400> 40
ggagagcatc 10

<210> 41
<211> 10
<212> DNA
<213> Homo sapien

<400> 41
gggagcaggg 10

<210> 42
<211> 10
<212> DNA
<213> Homo sapien

<400> 42
ctgagggtac 10

<210> 43
<211> 10
<212> DNA
<213> Homo sapien

<400> 43
aagctaccca 10

<210> 44
<211> 10
<212> DNA
<213> Homo sapien

<400> 44
cacccaggac 10

<210> 45
<211> 10
<212> DNA
<213> Homo sapien

<400> 45
acccaccctc 10

<210> 46
<211> 10
<212> DNA
<213> Homo sapien

<400> 46
accaacctga 10

<210> 47
<211> 10
<212> DNA
<213> Homo sapien

<400> 47
atgttgaccc 10

<210> 48
<211> 10
<212> DNA
<213> Homo sapien

<400> 48
cccagcgctg 10

<210> 49
<211> 5505
<212> DNA
<213> Homo sapien

<220>
<221> allele

<222> (1489)
<223> PS1: Polymorphic base C or T

<220>
<221> allele
<222> (1671)
<223> PS2: Polymorphic base G or A

<220>
<221> allele
<222> (1760)
<223> PS3: Polymorphic base C or T

<220>
<221> allele
<222> (1946)
<223> PS4: Polymorphic base C or T

<220>
<221> allele
<222> (2397)
<223> PS5: Polymorphic base A or G

<220>
<221> allele
<222> (2482)
<223> PS6: Polymorphic base G or T

<220>
<221> allele
<222> (2626)
<223> PS7: Polymorphic base G or A

<220>
<221> allele
<222> (2753)
<223> PS8: Polymorphic base C or T

<220>
<221> allele
<222> (3115)
<223> PS9: Polymorphic base G or A

<400> 49
gaattcgtaa ggttgaggag agttgatgat gccaagtact gtggtggtcc tgaacaatta 60
aagagggatt ctgggaagca gagtgtggac agttccaact ccctgccaag gggaagctca 120
taggcaaagg aagctcactc cagaggggat atggaagttc cataccctct tttgtctgaa 180
gagccgaagt ccctgttctc aggtcgttag gaagttaaaa agtaatttgg aggttatcag 240

aactgattga attgagtttg aacctcacct atagcaacaa tgggccaggc tgcttgacta 300
atgccttggc gtcaatggta cagttttctc cctcttgagc tgctggcagg gacctgggct 360
gacatgtctc agaaggccct tagtcaatga ttagcttatc tcaaggcccc aagccagggc 420
agctgtcaaa gagggtcccc actgcgttct gcacctagat cctcattgtg aaatgaagta 480
tgaagtgatt gagtgaggtc tctattgtct ctgacatttt acaattccag gattctgcct 540
tcttgtcaga gaagtgtata ggcaagcagt tgggcaggtg agagggctcc gggtgagagg 600
ctcagagact gagggctcag cctctgcttg aagagtcatc atctgggagg ctctgtggcc 660
tcctcagatg agttcattca cactcatacc ccaaaatgga gtcaacaccc cctccaccat 720
ctcagccttc tcagcatcta aagccccagc atcgatgcct ctttttttgg gttaggggtc 780
agagctgttg tggaagggca tacagtcatt cttcacttgc ctttgactgt gtactctgtg 840
cacatggagg taggagcaga catgacttca acaaggtcat gccccttgg caagcatctt 900
tgagaccaga gaggaagaca gactagggaa agaatgagga gataagcacg ggctgctgtg 960
aggtccaggg gagcaggcaa aggtaagaga aaaggcttta ggatactaac taacatatat 1020
ggagcactag catgagccag gcactattct aagtgctttt caggtgttat ctctttttgc 1080
ctcacggaca gcacctacaa ggcactgtaa ttatccctac ttcacagatg agggagtgga 1140
gccacagtga ggttaactta cttgaccaag ggggccaagt aggaatggag gcatttgttg 1200
agtcttctaa agatgaggaa agagtggaag tgagattttg taagtgcttg attcatttct 1260
accaactgaa ctggcaaata aataaaagca tgagtaaatg ggggtataaa tagtctgtca 1320
gctatggggg tgggagtggg ctcaaggcag gcttagagag aaggtgcaag agctgtctga 1380
aaaggtcaga gcaaagcatg aagctggtga gcagctgtga ccatagctgg aagcttctct 1440
ctgagctttc tcctggttac ctcctcctcc cctacgtgac cagtcagcya agtgttaagt 1500
ccaggggaac attttgctgc ttccaagtac tgtctcacta gtgttatttg ccataacttg 1560
cggccacagg gcaaggtcca ggtgctcaga cctttacatc ctggactttc caaggcctcc 1620
caaagctctc tggcacccag ggaacagtgt gcgtgtcgag agcttaatcc rcaggagcat 1680
agccatggtt ctctggggtc cagtgctggg agctctgctg gtggtcattg ctggatacct 1740
gtgcctgcca gggatgctcy gacaacgcag gccatgggag ccccctctgg acaagggtac 1800
cgtgccctgg cttggccatg ccatggcttt ccggaagaat atgtttgaat ttctgaagcg 1860
catgaggacc aagcatgggg atgtgttcac agtgcagcta gggggccagt acttcacctt 1920
cgtcatggac cccctctcct ttggcyccat cctcaaggac acacagagaa aactagactt 1980
tgggcaatat gcaaaaaaac tggtgctgaa ggtatttgga taccgttcag tgcaagggga 2040
ccatgagatg atacactcag ccagcaccaa gcatctgagg ggggatggct tgaaggatct 2100
taatgagacc atgctggaca gcctgtcctt tgtaatgctg acgtccaaag gctggagtct 2160
ggatgccagt tgctggcatg aggacagcct ctttcgcttc tgctattaca tcttgttcac 2220
agctggctac ctgagcttgt tcggctacac gaaggacaag gagcaggacc tgctacaggc 2280
aggagagtta ttcatggagt tccgcaagtt tgaccttctt ttcccaaggt ttgtctactc 2340
cctgctgtgg ccccgggagt ggctagaagt gggccgactc cagcgtctct ttcacargat 2400
gctctccgtg agccacagcc aggagaagga gggcatcagc aactggctgg gcaacatgct 2460
tcagtttctg agggagcagg gkgtaccctc agctatgcag gacaagttca acttcatgat 2520
gctctgggcc tcccagggga acacggggcc tacctctttc tgggccctct tgtacctcct 2580
gaagcaccca gaagctattc gggctgtgag ggaggaagct acccargtcc tgggtgaggc 2640
caggctggag accaagcagt cctttgcctt caaactcggt gccctgcaac acaccccagt 2700
tctagacagc gtggtggagg agacgctgcg gctgagggct gcacccaccc tcytcaggtt 2760
ggttcatgaa gactataccc tgaagatgtc cagtgggcag gagtatctgt tccgccatgg 2820
agacatcctg gccctctttc cctacctctc agtgcacatg gaccctgaca tccaccctga 2880
gcccaccgtc ttcaagtacg atcgcttcct caaccctaat ggcagccgga aagtggactt 2940
cttcaagaca ggcaagaaga tccaccacta caccatgccc tggggttcgg gcgtttccat 3000
ctgccctggg aggttctttg cactcagtga ggtgaagctc tttatcctgc ttatggtcac 3060
acactttgac ttagagttgg tggaccctga cacaccacta ccccatgttg acccrcagcg 3120

ctggggtttt ggcaccatgc agcccagcca cgatgtgcgc ttccgctacc gcctgcatcc 3180
tacagagtga gcttggccaa gccagctgca aacctggcca gaggagttct attgcatctc 3240
tcacctgttc tcacccctct gcagccccaa gaccccactg gccacccctc cctctggtcc 3300
tgtggcaccc cctacctctg ttctgcctgt cctcgctctc tccccgccta gtcatctgac 3360
aggcttatca ttctctttaa aataccatct ctcagagtgg gttctgccga accctcctct 3420
cacaggaagt ccagaggaag ggggagtatc tgtgggcaac ttggtttggg agatgatgcc 3480
tgccttgaga agtcctgagt acagagactg gttcccccca gacacgagta acatggcatc 3540
ttgcaaacat cagcctccac tctcccagct tgctttagtt ttttcagcaa cacttatccc 3600
acatcctatg gaattcaggt tctagaacag tgtcatccaa cataaatatg aagcaagcta 3660
catgagtagt ggggtttttt ggttttgttt gtttgttttg agacagagtc ttgctctgtc 3720
gccgaggctg cagtgcagcg gtgcgatctc tgctcactgc aacctctgcc tcctgggttt 3780
aagcaattct cctgcctcag cctccccagt agctgggatt acaggcacct accaccttgc 3840
ccagctaatt tagtttcctg gtaaacacat ttttaaaaag taaaatgaaa caattcattt 3900
ttataatata ctttacttaa tccaatatat ccaaaatatt ggcatttcaa cgtgatcact 3960
attaaaaatt ttaacgtaat agtttttacc ttccctttt catcctgtct tcaaagtctg 4020
gtatgtactg tacttttgca gttcttccca gttcagacta gccacattcc aagtgcttaa 4080
ttgccgtgtg tggcaggtgg ctgccctatt ggatacaaca ggtctagaga aatgatacct 4140
ttttttttt tttgagacag agtctcactc tgttgcccag gctgcagtgc agtggtgtga 4200
tcttggctca ctacaacctc tgcctcccgg gttcaagtaa ttctcctgcc tcaggcttct 4260
gagtagctgg gattataggt gcgcaccacc acacccggct agtttctgta tttttagtag 4320
agacggcgtt tcaccatttt ggccggcctc ctcggcctcc caaagttctg ggattacagg 4380
catgagccac tgtgcccagt caggatgtct ttaatgtagg aatcattcaa gatccctcct 4440
cagtgcccat gtctccccca cctcagggtg ctgtcaaccc tcccttggtt tccatgatca 4500
ttgctgaact cagagctttc ttctatcccc agacccacat gggagtctcc cagcacctct 4560
gactcagctt ggccaatgca gaacttggca tctgaacccc tgcctcttcc ctgaaccact 4620
cctcctcctg atttccctgc tctctctgct cagaacacta ccattcaccc catctcccag 4680
agacactctc cctcctcctc tttcaccttc atacaagcag gcccccgtcc tatcagtaca 4740
ccccccctgtg atgcctctga aatatgtacc ctcctccct ctcctactgt cattcttata 4800
cctgtgatct ctgcgtttct cttctacact gctgccagag gccttttttgc caaaagcact 4860
gctgattgtc actcttttac tttcctggtt ccccctttgac ttcacacact tgacgtttac 4920
attttgtcac agttggtccc ccttgccttc acatacctga tgtttaaatt ttgtcacggt 4980
tggtctccaa acttctcaac tttgccagct ttctccaatg aaaccatcag tagagacatt 5040
gctgttcctg cccatcccag tgccagctac tccttcctgg aagccttcct ggatttcacg 5100
catcccatcc tctggccaga aatggcctcc gaactgtgca gctcctgcgg caccatttgt 5160
atcatagcat atctgttatg tgtcgttgct ttctgtgtct tggtttctgc ctgctgccca 5220
cttccctaga ctggagccta ctggggcatg gtattttatt catttctgac ctccaattcc 5280
ctaagtggag attaaatgtt agaaacctga cttccagtct cggctaattt actttcagtc 5340
ctcaaatcat ttcctggcct cagccctcca atctatgaaa tgaggacaat ccccccttccc 5400
agcgtgctta tcagtgctac aagagggtga gggggtggct gccatagctg cagctggctg 5460
atggcagggc tccagctgga gctgggacag gaggaagaat gactg 5505

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 October 2001 (25.10.2001)

PCT

(10) International Publication Number
WO 01/79224 A3

(51) International Patent Classification[7]: C12P 19/34, C12Q 1/68, C07H 21/04

(21) International Application Number: PCT/US01/11946

(22) International Filing Date: 12 April 2001 (12.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/196,408 12 April 2000 (12.04.2000) US

(71) Applicant *(for all designated States except US)*: GENAISSANCE PHARMACEUTICALS, INC. [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BENTIVEGNA, Steven, C. [US/US]; 49 Greenview Terrace, Middletown, CT 06457 (US). CHEW, Anne [US/US]; 1477 Beacon Street #64, Brookline, MA 02446 (US). CHOI, Julie, Y. [US/US]; 38 Elizabeth Street, West Haven, CT 06516 (US). KOSHY, Beena [IN/US]; Apt. 11B, 1298 Hartford Turnpike, North Haven, CT 06473 (US). STEPHENS, J., Claiborne [US/US]; 46 Crabapple Lane, Guilford, CT 06437 (US).

(74) Agent: FIELD, Gisela, M.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
28 February 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/79224 A3

(54) Title: HAPLOTYPES OF THE CYP8B1 GENE

(57) Abstract: Novel single nucleotide polymorphisms in the human cytochrome P450 subfamily VIIIB (CYP8B1) gene are described. In addition, various genotypes, haplotypes and haplotype pairs for the CYP8B1 gene that exist in the population are described. Compositions and methods for haplotyping and/or genotyping the CYP8B1 gene in an individual are also disclosed. Polynucleotides containing one or more of the CYP8B1 polymorphisms disclosed herein are also described.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/11946

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) C12P 19/34; C12Q 1/68; C07H 21/04
US CL 435/6, 91.2; 536/23.1, 24.3
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6, 91.2; 536/23.1, 24.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
A	GAFVELS, M. ET AL. Structure and Chromosomal Assignment of the Sterol 12.alpha.-Hydroxylase Gene (CYP8B1) in Human and Mouse: Eukaryotic Cytochrome P-450 Gene Devoid of Introns. Genomics. 1999. Vol. 56, pages 184-196.	1-3, 7, 8

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 July 2001 (23.07.2001)	19 NOV 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Juliet C. Einsmann Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/11946

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: 1-3, 7, 9, all partially as they pertain to haplotypes 2 and 3

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☒ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No
PCT/US01/11946

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Groups 1-25, claim(s) 1-3, 7-8 in part, drawn to methods for haplotyping CYP8B1 comprising determining whether the individual has one of the CYP8B1 haplotypes shown in Table 5 or one of the haplotype pairs shown in Table 4. It is noted that Groups 1-25 correspond to the haplotypes of Table 5 and the haplotype pairs of Table 4, respectively. For example if Group 1 is elected, the claims 1-3 and 7-8 will be examined to the extent that they apply to methods of haplotyping comprising a step of determining whether the individual has the first haplotype of Table 5 of the CYP8B1 gene. Upon election of an invention in this group, please specify the Table and number of haplotypes requested.

Groups 26-34, claim(s) 4-6, in part drawn to a method for genotyping the CYP8B1 gene. It is noted that Groups 118-140 correspond to polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, and PS9, respectively. For example, if Group 118 is elected, the claims 4-6 will be examined to the extent that they apply are limited to method of genotyping comprising a step of identifying the nucleotide pair at PS1.

Groups 35-70, claim(s) 9-10, in part drawn to a method for predicting a haplotype pair for the CYP8B1 gene by identifying a CYP8B1 genotype for the individual at two or more polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, and PS9. It is noted that the claims encompass methods requiring 35-70 each correspond to one of these possible pairs, in the order recited in the claim. For example, if Group 35 is selected, the claim 9-10 will be examined to the extent that it applies to a combination of PS1 and PS2. If Group 70 is selected, the claim 10 will be examined to the extent that it applies to a combination of PS8 and PS9. If applicants elect any of these groups, please specify the two sites to be examined in the method for predicting a haplotype.

Groups 71-95, claim(s) 11-12, in part drawn to a method for identifying an association between a trait and a haplotype between one of the 25 haplotypes and haplotype pairs of CYP8B1 gene. Groups 71-95 each correspond to one of the 25 particular combinations of the polymorphic sites, haplotypes, and the haplotype pairs encompassed by the claims (i.e., the 12 different haplotypes of Table 5, as well as the 13 different haplotype pairs of Table 4). For example if Group 71 is selected, the claims will be examined to the extent that they apply to the first haplotype of Table 5.

Groups 96-104, claim(s) 13-17, in part, drawn to a composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the CYP8B1 gene.

Group 105, claims 18, drawn to a kit comprising a set of oligonucleotides designed to genotype each of the polymorphic sites.

Groups 106-119, claims 19-20 and 23-24, in part, drawn to a polynucleotide which is a polymorphic variant of a reference sequence for CYP8B1 gene, or a fragment thereof.

Group 120-131, claim(s) 21-22 and 25-26, in part drawn to a recombinant nonhuman organisms comprising one of the 12 haplotypes respectively. If Group 120 is selected the transgenic organism will be examined to the extent that it applies to haplotype 1.

Group 132-143, claim(s) 27, in part drawn to a polypeptide comprising an amino acid sequence which is a polymorphic variants of a reference sequence for the CYP8B1 protein or a fragment thereof.

Group 144-155, claim(s) 28, in part drawn to an antibody which binds to a polypeptides of Claim 29.

Group 156-164, claim(s) 29, in part drawn to a method for screening for drugs targeting the CYP8B1 polypeptide.

Group 165-189, claim(s) 30, in part drawn to a computer system comprising polymorphism data wherein the data comprises the haplotypes shown in Table 5 and the haplotype pairs of Table 4. Group 165 is selected, the computer system will be examined to the extent that it applies to the first of 12 haplotypes.

Groups 190-201, claim(s) 31, in part, drawn to a genome anthologies comprising CYP8B1 isogenes having any one of the haplotypes of Table 5. It is noted that Groups 190-201 correspond to anthologies comprising one of the haplotypes 1-12 of Table 5 in the order shown in Table 5. For example, Group 190 is drawn to an anthology comprising haplotype 1 of the Table 5.

The products claimed in Claims 13-20 and 23-24 include fragments of variant sequences, and the claims do not require, e.g., that the recited polymorphic sites be included in said fragments. Accordingly, the claims are sufficiently broad so as to encompass nucleic acid fragments of the CYP8B1 gene. The description teaches that this gene was known in the prior art, and as such, fragments of this gene are obvious over the disclosure of the full length. As the products encompassed by these claims do not represent a contribution over the prior art, the claims lack a special technical feature that is the same as or that corresponds to a special technical feature of the other claimed inventions. Thus, there is no special technical feature linking the recited Groups, as would be necessary to fulfill the requirement for unity of invention.

It is also noted that each of the present claims has been presented in improper Markush format, as distinct products and distinct methods are improperly joined in the claims. Each polymorphic site and each molecule containing said polymorphic site is structurally and functionally distinct from and has a different special technical feature than each other polymorphic site and molecules containing said site. The chemical structure of each polymorphism and of each molecule containing the same differ from each other. For example, a polynucleotide comprising PS1 is chemically, structurally, and functionally different from a molecule comprising PS4. As the products and methods encompassed by the claims do not share a special technical feature, the distinct products and methods may not properly be presented in the alternative. Accordingly, the claims have been separated into a number of groups corresponding to the number of different inventions encompassed by the claims, and the claims will be examined only as they read upon the invention of the elected group. For the same reasons, the remainder of the claims have been separated in a number of groups corresponding to the number of different inventions encompassed thereby.

It is noted that the haplotypes and genotypes encompassed by the instantly recited method claims are also distinct from each other and from the single polymorphisms recited in e.g., claims 4-6. For example, a molecule of haplotype 1, comprising a particular combination of polymorphisms, differs chemically, structurally, and functionally from a molecule of haplotype 2 and from a molecule comprising a single polymorphism (e.g., PS1). The special technical feature of each haplotype or genotype is the combination of polymorphisms contained therein, which feature is lacking from and not shared with each other haplotype or genotype or with, e.g., a molecule comprising any single polymorphism set forth in the claims. Similarly, with respect to the pairs of polymorphisms of Claim 9, each combination of polymorphisms differs from each other combination and from each of the other combinations discussed above (i.e., haplotypes, genotypes, and single polymorphic sites). Thus, the claims have been separated into a number of groups corresponding to the number of different inventions encompassed thereby, and the claims will be examined only as they read upon the invention of the elected group.

Further the groups comprising polynucleotides, kits, recombinant organisms, polypeptides, antibodies, computer systems and genome anthologies are additionally drawn to multiple, distinct products lacking the same or corresponding special technical features. The nucleic acids are composed of nucleotides and function in, e.g., methods of nucleic acid hybridization or amplification. These groups are directed to different combinations of nucleic acids which are different from one another and may be employed in different methods. The recombinant organisms are complex organisms that are employed in, e.g. animal research methods. Such organisms cannot be employed as, e.g., probes or primers and they differ in both structure and function from the nucleic acids. The polypeptides differ in both structure and function from either the nucleic acids or the transgenic organisms. The polypeptides are composed of amino acids linked by peptide bonds and arranged in a complex combination of alpha helices, beta pleated sheets, hydrophobic and hydrophilic domains. The polypeptides also differ in function, e.g., fusion proteins with an enzymatic functions. The antibodies are composed of amino acids linked by peptide bonds, antibodies are glycosylated and their tertiary structure is unique, where the subunits (2 light chains and 2 heavy chains) associated via disulfide bonds into a Y-shaped symmetric dimer. The antibodies function in immunoassays. Further the computer systems are composed of, e.g., a CPU, a display device, an input device, etc., as recited in Claim 30, and function in, e.g., methods of electronic sequence comparison. Accordingly, the products differ structurally and functionally from one another. As products of different sets of Groups differ from each other in structure, function, and effect, they do not belong to a recognized class of chemical compound, or have both a "common property or activity" and a common structure as would be required to show that the inventions are "of a similar nature".

Further, the different methods have different objectives and require different process steps. The haplotyping methods require steps of identifying haplotypes and haplotype pairs to achieve the objectives of haplotyping. The methods of genotyping require steps of identifying a single nucleotide on one gene copy to achieve the objective of genotyping. The methods of predicting a haplotype pair require steps of identifying two polymorphisms in a gene to achieve the objective of "predicting a haplotype pair". The methods of identifying an association requires steps of comparing frequencies of haplotypes in a population to achieve the objective of "identifying an association between a trait" and a haplotype. The methods of assaying for binding activity require steps of assaying for binding activity for candidate agents. In addition to differences in objectives, effects, and method steps, it is again noted that the claims of the present Groups are not directed to the detection or identification of molecules having the same or common special technical feature, for the reasons discussed above.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/11946

Continuation of B. FIELDS SEARCHED Item 3:
US PATENTS DATABASE, BIOSIS, MEDLINE, CAPLUS
search terms: cytochrome p450 VIIIB, cyp8b1, sterol 12-alpha-hydroxylase, polymorphism, mutation, variant, allele
searched SEQ ID NO: 1-12 in EMBL, GenBank, US Patents, Genseq, and EST databases